

GRAPHICAL MODELING OF SPATIAL HEALTH DATA

ADRIAN DOBRA

ABSTRACT. The literature on Gaussian graphical models (GGMs) contains two equally rich and equally significant domains of research efforts and interests. The first research domain relates to the problem of graph determination. That is, the underlying graph is unknown and needs to be inferred from the data. The second research domain dominates the applications in spatial epidemiology. In this context GGMs are typically referred to as Gaussian Markov random fields (GMRFs). Here the underlying graph is assumed to be known: the vertices correspond to geographical areas, while the edges are associated with areas that are considered to be neighbors of each other (e.g., if they share a border). We introduce multi-way Gaussian graphical models that unify the statistical approaches to inference for spatiotemporal epidemiology with the literature on general GGMs. The novelty of the proposed work consists of the addition of the G-Wishart distribution to the substantial collection of statistical tools used to model multivariate areal data. As opposed to fixed graphs that describe geography, there is an inherent uncertainty related to graph determination across the other dimensions of the data. Our new class of methods for spatial epidemiology allow the simultaneous use of GGMs to represent known spatial dependencies and to determine unknown dependencies in the other dimensions of the data.

KEYWORDS: Gaussian graphical models, Gaussian Markov random fields, spatiotemporal multivariate models

CONTENTS

1. Introduction	1
2. GGMs vs. GMRFs	2
3. Multi-way Gaussian Graphical Models	3
4. Inference in Multi-way GGMs	5
5. Multi-way GGMs with Separable Mean Parameters	6
6. An Application to Spatiotemporal Cancer Mortality Surveillance	7
7. Dynamic Multi-way GGMs for Array-variate Time Series	9
8. Discussion	11
References	12

1. INTRODUCTION

Graphical models [62, 40] that encode multivariate independence and conditional independence relationships among observed variables $X = (X_1, \dots, X_p)$ have a widespread use in major scientific areas (e.g., biomedical and social sciences). In particular, a Gaussian graphical model (GGM) is obtained by setting off-diagonal elements of the precision matrix $K = \Sigma^{-1}$ to zero of a p -dimensional multivariate normal model [22]. Employing a GGM instead of a multivariate normal model leads to a significant reduction in the number of parameters that need to be estimated if most elements of K are constrained to be zero and p is large. A pattern of zero constraints in K can be recorded as an undirected graph $G = (V, E)$ where the set of vertices $V = \{1, 2, \dots, p\}$ represent observed variables, while the set of edges $E \subset V \times V$ link all the pairs of vertices that correspond to off-diagonal elements of K that have not been set to zero. The absence of an edge between X_{v_1} and X_{v_2} corresponds with the conditional independence of these two random variables given the rest and is denoted by $X_{v_1} \perp\!\!\!\perp X_{v_2} \mid X_{V \setminus \{v_1, v_2\}}$ [60]. This is called the pairwise Markov property relative to G , which in turn implies the local as well as the global Markov properties relative to G [40]. The local Markov property plays a key role since it gives the regression model induced by G on each variable X_v . More explicitly, consider the neighbors of v in G , that is, the set of vertices $v' \in V$ such that $(v, v') \in E$. We denote this set by $\text{bd}_G(v)$. The local Markov property

relative to G says that $X_v \perp\!\!\!\perp X_{V \setminus \{v\} \cup \text{bd}_G(v)} \mid X_{\text{bd}_G(v)}$. This statement is precisely the statement we make when we drop the variables $\{X_{v'} : v' \in V \setminus \text{bd}_G(v)\}$ from the regression of X_v on $\{X_{v'} : v' \in V \setminus \{v\}\}$.

The literature on GGMs contains two equally rich and significant domains of research. The first research domain relates to the problem of graph determination. That is, the underlying graph is unknown and needs to be inferred from the data. Frequentist methods estimate K and Σ given one graph that is best supported by the data in the presence of sparsity constraints that penalize for increased model complexity (i.e., for the addition of extra edges in the graph). Among numerous notable contributions we mention the regularization methods of [47, 65, 10, 29] as well as the simultaneous confidence intervals of [27]. Bayesian methods proceed by imposing suitable prior distributions for K or Σ [43, 64, 18, 5, 53, 45, 49]. Inference can be performed based on the best model, i.e. the graph having the highest posterior probability, or by Bayesian model averaging [38] over all $2^{p(p-1)/2}$ possible graphs using Markov chain Monte Carlo (MCMC) approaches [34, 21, 63]. As the number of graphs grows, MCMC methods are likely to visit only subsets of graphs that have high posterior probabilities. To this end, various papers [37, 52, 42] have proposed stochastic search methods for fast identification of these high posterior probability graphs.

The second research domain on GGMs dominates the applications in spatial epidemiology. In this context GGMs are referred to as Gaussian Markov random fields (GMRFs) [6, 7, 8]. The underlying graph G is assumed to be known: the vertices correspond to geographical areas, while the edges are associated with areas that are considered to be neighbors of each other (e.g., if they share a border). A GMRF is specified through the conditional distributions of each variable given the rest

$$(1) \quad \{p(X_v \mid X_{V \setminus \{v\}}) : v \in V\},$$

which are assumed to be normal. The local Markov property leads to a further reduction in the set of full conditionals:

$$(2) \quad p(X_v \mid X_{V \setminus \{v\}}) = p(X_v \mid X_{\text{bd}_G(v)}).$$

Since it is typically assumed that phenomena (e.g., the occurrence of a disease) taking place in one area influence corresponding phenomena taking place in the remaining areas only through neighbor areas, the set of reduced conditionals (2) are employed to describe a full joint distribution of random spatial effects. GMRFs are conditional autoregressions (CAR) models that have a subclass called simultaneous autoregressions (SAR). For a comprehensive account of inference in CAR/SAR/GMRFs see [17, 51, 32]. Key questions relate to conditions in which a joint distribution determined by (1) actually exists and, if it does, whether it is multivariate normal. This leads to particular parametric specifications for the set of conditionals (1) and (2) that are more restrictive than the general parametric specification of a GGM.

In this chapter we examine the theoretical differences between GGMs and GMRFs. [26] developed efficient MCMC methods for inference in univariate and matrix-variate GGMs, and subsequently employed these methods to construct Bayesian hierarchical spatial models for mapping multiple diseases. We extend their results to multi-way GGMs that can capture temporal dependencies in addition to several other relevant dimensions. We exemplify the construction of a Bayesian hierarchical spatiotemporal model based on three-way GGMs, and also present a related theoretical extension of multi-way GGMs to dynamic multi-way GGMs for array-variate time series.

2. GGMs vs. GMRFs

We consider a GGM defined by a graph $G = (V, E)$ for the multivariate normal distribution $N_p(0, K^{-1})$ of a vector $X = (X_1, \dots, X_p)$. The precision matrix K is constrained to belong to the cone P_G of positive definite matrices such that $K_{ij} = 0$ for all $(i, j) \notin E$. The full conditionals (1) associated with each X_v are expressed as a function of the elements of K as follows:

$$(3) \quad X_v \mid X_{V \setminus \{v\}} = x_{V \setminus \{v\}} \sim N\left(- \sum_{v' \in \text{bd}_G(v)} (K_{vv'} / K_{vv}) x_{v'}, 1 / K_{vv}\right).$$

Remark that the variables $X_{v'}$ that are not linked by an edge with X_v are dropped from the full conditional (3) because $K_{vv'} = 0$. A GMRF with graph G is parametrized through the full conditionals

$$(4) \quad X_v \mid X_{V \setminus \{v\}} = x_{V \setminus \{v\}} \sim N\left(\sum_{v' \in \text{bd}_G(v)} \beta_{vv'} x_{v'}, \sigma_v^2\right), \text{ for } v \in V.$$

The symmetry condition $\beta_{vv'}\sigma_v^2 = \beta_{v'v}\sigma_{v'}^2$ for all $v \neq v'$ is necessary for the conditionals (4) to define a proper GGM [40]. Additional constraints under which the set of regression parameters $\{(\beta_{vv'}, \sigma_v^2)\}$ induce a proper precision matrix $K \in \mathbf{P}_G$ with $K_{vv} = \sigma_v^{-2}$ and $K_{vv'} = -\beta_{vv'}\sigma_v^{-2}$ are given in [8]. [9] make use of a symmetric proximity matrix W with $w_{vv} = 0$, $w_{vv'} > 0$ if $v' \in \text{bd}_G(v)$ and $w_{vv'} = 0$ if $v' \notin \text{bd}_G(v)$. They define $\beta_{vv'} = \rho w_{vv'}/w_{v+}$ and $\sigma_v^2 = \sigma^2/w_{v+}$, where $w_{v+} = \sum_{v'} w_{vv'}$. Here ρ is referred to as a spatial autocorrelation parameter. With this choice, for each $\rho \in (-1, 1)$ and $\sigma^2 > 0$, the GMRF specified by the full conditionals (4) has a precision $K = \sigma^2(E_W - \rho W)^{-1} \in \mathbf{P}_G$, where $E_W = \text{diag}\{w_{1+}, \dots, w_{p+}\}$. As such, this widely used parametrization of GMRFs is quite restrictive since not any matrix in the cone \mathbf{P}_G can be represented through the two parameters ρ and σ^2 given a particular choice of W . This difficulty originates from the parametrization (4) of a GGM. Instead, by imposing proper prior distributions for the precision matrix K , we avoid the unnecessary representation of a GGM as the set of full conditionals (4). In particular, we use of the G-Wishart prior $\text{Wis}_G(\delta, D)$ with density

$$(5) \quad p(K | G, \delta, D) = \frac{1}{I_G(\delta, D)} (\det K)^{(\delta-2)/2} \exp\{-\frac{1}{2}\langle K, D \rangle\},$$

with respect to the Lebesgue measure on \mathbf{P}_G [50, 4, 44]. Here $\langle A, B \rangle = \text{tr}(A^T B)$ denotes the trace inner product. The normalizing constant $I_G(\delta, D)$ is finite provided $\delta > 2$ and D positive definite [23]. The G-Wishart prior $\text{Wis}_G(\delta, D)$ is conjugate to the normal likelihood. For a thorough account of its numerical properties see [42] and the references therein.

For applications in hierarchical spatial models, [26] set $D = (\delta - 2)\sigma^2(E_W - \rho W)^{-1}$ because, with this choice, the prior mode for K is precisely $\sigma^2(E_W - \rho W)^{-1}$ – the precision matrix of a GMRF. The prior specification for the precision matrix can therefore be completed in a manner similar to the current work from the existent literature on GMRFs. We note that the G-Wishart prior for K induces compatible prior distributions for the regression parameters (4) – see [24]. The advantage of this representation of GGMs is a more flexible framework for GMRFs that allows their regression coefficients to be determined from the data rather than being fixed or allowed to vary as a function of only two parameters.

3. MULTI-WAY GAUSSIAN GRAPHICAL MODELS

We develop a framework for analyzing datasets that are associated with a random L -dimensional array \mathbf{X} . Such datasets are quite common in social and biomedical sciences. In particular, spatial epidemiology involves datasets recording SIRs of several diseases observed under different conditions at multiple time points. The notations, definitions and operators related to tensors that appear throughout are introduced in [20, 39]. The elements of the observed multi-way array are indexed by $\{(i_1, i_2, \dots, i_L) : 1 \leq i_l \leq m_l\}$. The total number of elements of \mathbf{X} is $m = \prod_{l=1}^L m_l$. We assume that \mathbf{X} follows an array normal distribution

$$\text{vec}(\mathbf{X}) | \mathbf{K} \sim \mathbf{N}_m(0, \mathbf{K}^{-1}),$$

whose $m \times m$ precision matrix \mathbf{K} is separable across each dimension, i.e.

$$(6) \quad \mathbf{K} = \mathbf{K}_L \otimes \mathbf{K}_{L-1} \otimes \dots \otimes \mathbf{K}_1.$$

The $m_l \times m_l$ precision matrix \mathbf{K}_l is associated with dimension l , while $\text{vec}(\mathbf{X})$ is the vectorized version of \mathbf{X} . The separability assumption might seem restrictive in the sense that it captures only dependencies across each dimension of the data without directly taking into account the interactions that might exist among two, three or more dimensions. However, this assumption reduces the number of parameters of the distribution of $\text{vec}(\mathbf{X})$ from $2^{-1}m(m+1)$ to $2^{-L} \prod_{l=1}^L m_l(m_l+1)$ which constitutes a substantial advantage when a sample size is small. The probability density of \mathbf{X} as an array is (see [36])

$$(7) \quad p(\mathbf{X} | \mathbf{K}_1, \dots, \mathbf{K}_L) = (2\pi)^{-\frac{m}{2}} \left[\prod_{l=1}^L (\det \mathbf{K}_l)^{\frac{1}{m_l}} \right]^{\frac{m}{2}} \exp \left\{ -\frac{1}{2} \langle \mathbf{X}, \mathbf{X} \times \{\mathbf{K}_1, \dots, \mathbf{K}_L\} \rangle \right\},$$

where

$$\mathbf{X} \times \{\mathbf{K}_1, \dots, \mathbf{K}_L\} = \mathbf{X} \times_1 \mathbf{K}_1 \times_2 \dots \times_L \mathbf{K}_L,$$

is the Tucker product. Here $\mathbf{X} \times_l \mathbf{K}_l$ is the l -mode product of the tensor \mathbf{X} and matrix \mathbf{K}_l . We refer to (7) as the mean-zero L -dimensional array normal distribution $\text{AN}_L(\mathbf{0}; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\})$.

Most of the existent literature has focused on two-dimensional (or matrix-variate) arrays – see, for example, [3, 48] and the references therein. [31] studies the separable normal model (6) for $L = 3$, while [19] presents theoretical results for matrix-variate distributions that includes (6) with $L = 2$ as a particular case. [36] has proposed a Bayesian inference framework for model (6) for an arbitrary number L of dimensions by assigning independent inverse-Wishart priors for the covariance matrices $\Sigma_l = \mathbf{K}_l^{-1}$ associated with each dimension. Despite its flexibility and generality, the framework of [36] does not allow any further reduction in the number of parameters of model (6). To this end, we propose a framework in which each precision matrix \mathbf{K}_l is constrained to belong to a cone P_{G_l} associated with a GGM with graph $G_l \in \mathcal{G}_{m_l}$. We denote by \mathcal{G}_{m_l} the set of undirected graphs with m_l vertices. Sparse graphs associated with each dimension lead to sparse precision matrices, hence the number of parameters that need to be estimated could be significantly smaller than $2^{-L} \prod_{l=1}^L m_l(m_l + 1)$. A similar framework has been proposed in [59] for matrix-variate data ($L = 2$) and for row and column graphs restricted to the class of decomposable graphs. Our framework is applicable for any number of dimensions and allows arbitrary graphs (decomposable and non-decomposable) to be associated with each precision matrix \mathbf{K}_l .

The prior specification for $\{\mathbf{K}_l\}_{l=1}^L$ must take into account the fact that two precision matrices are not uniquely identified from their Kronecker product which means that, for any $z > 0$ and $l_1 \neq l_2$,

$$\mathbf{K}_L \otimes \cdots \otimes (z^{-1} \mathbf{K}_{l_1}) \otimes \cdots \otimes (z \mathbf{K}_{l_2}) \otimes \cdots \otimes \mathbf{K}_1 = \mathbf{K}_L \otimes \cdots \otimes \mathbf{K}_{l_1} \otimes \cdots \otimes \mathbf{K}_{l_2} \otimes \cdots \otimes \mathbf{K}_1$$

represents the same precision matrix for $\text{vec}(\mathbf{X})$. We follow the basic idea laid out in [59] and impose the constraints

$$(8) \quad (\mathbf{K}_l)_{11} = 1, \text{ for } l = 2, \dots, L.$$

Furthermore, we define a prior for \mathbf{K}_l , $l \geq 2$, through parameter expansion by assuming a G-Wishart prior $\text{Wis}_{G_l}(\delta_l, \mathbf{D}_l)$ for the matrix $z_l \mathbf{K}_l$ with $z_l > 0$, $\delta_l > 2$ and $\mathbf{D}_l \in \mathbf{P}_{G_l}$. We denote $G_l = (V_l, E_l)$, where $V_l = \{1, 2, \dots, m_l\}$ are vertices and $E_l \subset V_l \times V_l$ are edges. We consider the Cholesky decompositions of the precision matrices from (6),

$$(9) \quad \mathbf{K}_l = \boldsymbol{\phi}_l^T \boldsymbol{\phi}_l,$$

where $\boldsymbol{\phi}_l$ is an upper triangular matrix with $(\boldsymbol{\phi}_l)_{ii} > 0$, $1 \leq i \leq m_l$. [50] proves that the set $\nu(G_l)$ of the free elements of $\boldsymbol{\phi}_l$ consists of the diagonal elements together with the elements that correspond with the edges of G_l , i.e.

$$\nu(G_l) = \{(i, i) : i \in V_l\} \cup \{(i, j) : i < j \text{ and } (i, j) \in E_l\}.$$

Once the free elements of $\boldsymbol{\phi}_l$ are known, the remaining elements are also known. Specifically, $(\boldsymbol{\phi}_l)_{1j} = 0$ if $j \geq 2$ and $(1, j) \notin E_l$. We also have

$$(\boldsymbol{\phi}_l)_{ij} = -\frac{1}{(\boldsymbol{\phi}_l)_{ii}} \sum_{k=1}^{i-1} (\boldsymbol{\phi}_l)_{ki} (\boldsymbol{\phi}_l)_{kj},$$

for $2 \leq i < j$ and $(i, j) \notin E_l$. The determination of the elements of $\boldsymbol{\phi}_l$ that are not free based on the elements of $\boldsymbol{\phi}_l$ that are free is called the completion of $\boldsymbol{\phi}_l$ with respect to G_l [50, 4]. It is useful to remark that the free elements of $\boldsymbol{\phi}_l$ fully determine the matrix \mathbf{K}_l . The development of our framework involves the Jacobian of the transformation that maps $\mathbf{K}_l \in \mathbf{P}_{G_l}$ to the free elements of $\boldsymbol{\phi}_l$ [50]:

$$J(\mathbf{K}_l \rightarrow \boldsymbol{\phi}_l) = 2^{m_l} \prod_{i=1}^{m_l} (\boldsymbol{\phi}_l)_{ii}^{d_i^{G_l} + 1},$$

where $d_i^{G_l}$ is the number of elements in $\text{bd}_{G_l}(i) \cap \{i + 1, \dots, m_l\}$ and $\text{bd}_{G_l}(i) = \{j : (i, j) \in E_l\}$ is the boundary of vertex i in G_l .

Our proposed prior specification for the separable normal model (7) is

$$(10) \quad \mathbf{K}_1 \mid \delta_1, \mathbf{D}_1 \sim \text{Wis}_{G_1}(\delta_1, \mathbf{D}_1), \quad (z_l \mathbf{K}_l) \mid \delta_l, \mathbf{D}_l \sim \text{Wis}_{G_l}(\delta_l, \mathbf{D}_l), \text{ for } l = 2, \dots, L.$$

The prior for \mathbf{K}_1 is

$$(11) \quad p(\mathbf{K}_1 \mid G_1) = \frac{1}{I_{G_1}(\delta_1, \mathbf{D}_1)} (\det \mathbf{K}_1)^{\frac{\delta_1 - 2}{2}} \exp \left\{ -\frac{1}{2} \langle \mathbf{K}_1, \mathbf{D}_1 \rangle \right\},$$

while the joint prior for (z_l, \mathbf{K}_l) is

$$(12) \quad p(z_l, \mathbf{K}_l | G_l) = \frac{1}{I_{G_l}(\delta_l, \mathbf{D}_l)} (\det \mathbf{K}_l)^{\frac{\delta_l-2}{2}} \exp \left\{ -\frac{1}{2} \langle \mathbf{K}_l, z_l \mathbf{D}_l \rangle \right\} z_l^{\frac{m_l(\delta_l-2)}{2} + |\nu(G_l)|-1},$$

for $l = 2, \dots, L$.

4. INFERENCE IN MULTI-WAY GGMs

We assume that the observed samples $\mathcal{D} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$ are independently generated from the mean-zero array normal distribution $\text{AN}_L(\mathbf{0}; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\})$. The resulting likelihood is expressed by introducing an additional dimension $m_{L+1} = n$ with precision matrix $\mathbf{K}_{L+1} = \mathbf{I}_n$, where \mathbf{I}_n is the $n \times n$ identity matrix. We see \mathcal{D} as a $m_1 \times \dots \times m_{L+1}$ array that follows an array normal distribution $\text{AN}_{L+1}(\mathbf{0}; \{m_1, \mathbf{K}_1\}, \dots, \{m_{L+1}, \mathbf{K}_{L+1}\})$. Furthermore, we define $\boldsymbol{\phi}_{L+1} = \mathbf{I}_n$. The Cholesky decompositions (9) of the precision matrices \mathbf{K}_l give the following form of the likelihood of \mathcal{D} :

$$(13) \quad p(\mathcal{D} | \mathbf{K}_1, \dots, \mathbf{K}_L) \propto \left[\prod_{l=1}^L (\det \mathbf{K}_l)^{\frac{1}{m_l}} \right]^{\frac{mn}{2}} \exp \left\{ -\frac{1}{2} \|\mathcal{D} \times \{\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_{L+1}\}\| \right\},$$

where $\|\mathbf{Y}\| = \langle \mathbf{Y}, \mathbf{Y} \rangle$ is the array norm [39]. Simple calculations show that the part of the likelihood (13) that depends on the precision matrix \mathbf{K}_l is written as:

$$(14) \quad p(\mathcal{D} | \mathbf{K}_1, \dots, \mathbf{K}_L) \propto (\det \mathbf{K}_l)^{\frac{mn}{2m_l}} \exp \left\{ -\frac{1}{2} \langle \mathbf{K}_l, \mathbf{S}_l \rangle \right\},$$

where $\mathbf{S}_l = \mathcal{D}_{(l)}^{[l]} \left(\mathcal{D}_{(l)}^{[l]} \right)^T$ and

$$\mathcal{D}_{(l)}^{[l]} = \mathcal{D} \times_1 \boldsymbol{\phi}_1 \times_2 \dots \times_{l-1} \boldsymbol{\phi}_{l-1} \times_l \mathbf{I}_l \times_{l+1} \boldsymbol{\phi}_{l+1} \times_{l+2} \dots \times_{L+1} \boldsymbol{\phi}_{L+1}.$$

Here $\mathbf{Y}_{(l)}$ is the l -mode matricization of an array \mathbf{Y} [39]. We develop a Markov chain Monte Carlo sampler from the posterior distribution of precision matrices $\mathbf{K}_l \in \mathbf{P}_{G_l}$, graphs $G_l \in \mathcal{G}_{m_l}$ and auxiliary variables z_l for $1 \leq l \leq L$:

$$(15) \quad p(\mathbf{K}_1, G_1, (\mathbf{K}_l, G_l, z_l)_{l=2}^L | \mathcal{D}) \propto p(\mathcal{D} | \mathbf{K}_1, \dots, \mathbf{K}_L) p(\mathbf{K}_1 | G_1) \prod_{l=2}^L p(z_l, \mathbf{K}_l | G_l) \prod_{l=1}^L \pi_{m_l}(G_l)$$

Here $\pi_{m_l}(G_l)$ are prior probabilities on the set of graphs \mathcal{G}_{m_l} . The full conditionals of \mathbf{K}_l , $1 \leq l \leq L$ and z_l , $2 \leq l \leq L$ are G -Wishart and Gamma, respectively:

$$\begin{aligned} p(\mathbf{K}_l | \text{rest}) &= \text{Wis}_{G_l} \left(\frac{mn}{m_l} + \delta_l, \mathbf{S}_l + z_l \mathbf{D}_l \right), \\ p(z_l | \text{rest}) &= \text{Gamma} \left(\frac{m_l(\delta_l - 2)}{2} + |\nu(G_l)|, \frac{1}{2} \langle \mathbf{K}_l, \mathbf{D}_l \rangle \right), \end{aligned}$$

where $\text{Gamma}(\alpha, \beta)$ has mean α/β . We use the approach for updating \mathbf{K}_l ($1 \leq l \leq L$) described in [25]. Their method sequentially perturbs each free element in the Cholesky decomposition of each precision matrix. The constraint (8) is imposed by not updating the free element $(\boldsymbol{\phi}_l)_{11} = \sqrt{(\mathbf{K}_l)_{11}} = 1$.

The updates of the graphs G_l are based on the full joint conditionals of \mathbf{K}_l and G_l , $1 \leq l \leq L$:

$$p(\mathbf{K}_l, G_l | \text{rest}) \propto \frac{1}{I_{G_l}(\delta_l, \mathbf{D}_l)} (\det \mathbf{K}_l)^{\frac{1}{2} \left(\frac{mn}{m_l} + \delta_l - 2 \right)} z_l^{\frac{m_l(\delta_l-2)}{2} + |\nu(G_l)|-1} \exp \left\{ -\frac{1}{2} \langle \mathbf{K}_l, \mathbf{S}_l + z_l \mathbf{D}_l \rangle \right\},$$

since, once an edge in G_l is added or deleted, the corresponding set of free elements of \mathbf{K}_l together with the remaining bound elements must also be updated.

We denote by $\text{nbd}_{m_l}^+(G_l)$ the graphs that can be obtained by adding an edge to a graph $G_l \in \mathcal{G}_{m_l}$ and by $\text{nbd}_{m_l}^-(G_l)$ the graphs that are obtained by deleting an edge from G_l . We call the one-edge-way set of graphs $\text{nbd}_{m_l}(G_l) = \text{nbd}_{m_l}^+(G_l) \cup \text{nbd}_{m_l}^-(G_l)$ the neighborhood of G_l in \mathcal{G}_{m_l} . These neighborhoods connect

any two graphs in \mathcal{G}_{m_l} through a sequence of graphs such that two consecutive graphs in this sequence are each others' neighbors. We sample a candidate graph $G'_l \in \text{nbd}_{m_l}(G_l)$ from the proposal distribution:

$$(16) \quad q(G'_l | G_l, z_l) = \frac{1}{2} \frac{z_l^{|\nu(G'_l)|}}{\sum_{G''_l \in \text{nbd}_{m_l}^+(G_l)} z_l^{|\nu(G''_l)|}} \delta_{\{G'_l \in \text{nbd}_{m_l}^+(G_l)\}} + \frac{1}{2} \frac{z_l^{|\nu(G'_l)|}}{\sum_{G''_l \in \text{nbd}_{m_l}^-(G_l)} z_l^{|\nu(G''_l)|}} \delta_{\{G'_l \in \text{nbd}_{m_l}^-(G_l)\}},$$

where δ_A is equal to 1 if A is true and is 0 otherwise. The proposal (16) gives an equal probability that the candidate graph is obtained by adding or deleting an edge from the current graph G_l .

We assume that the candidate graph G'_l is obtained by adding an edge (v_1, v_2) , $v_1 < v_2$, to G_l . We have $\nu(G'_l) = \nu(G_l) \cup \{(v_1, v_2)\}$, $\text{bd}_{G'_l}(v_1) = \text{bd}_{G_l}(v_1) \cup \{v_2\}$ and $d_{v_1}^{G'_l} = d_{v_1}^{G_l} + 1$. We define an upper diagonal matrix ϕ'_l such that $(\phi'_l)_{v'_1, v'_2} = (\phi_l)_{v'_1, v'_2}$ for all $(v'_1, v'_2) \in \nu(G_l)$. The value of $(\phi'_l)_{v_1, v_2}$ sampled from a $\mathcal{N}((\phi_l)_{v_1, v_2}, \sigma_g^2)$ distribution. The bound elements of ϕ'_l are determined through completion with respect to G'_l . We form the candidate matrix $\mathbf{K}'_l = (\phi'_l)^T \phi'_l \in \mathbf{P}_{G'_l}$. Since the dimensionality of the parameter space increases by one, we must make use of the reversible jump method of [35]. We accept the update of (\mathbf{K}_l, G_l) to (\mathbf{K}'_l, G'_l) with probability $\min\{R_g, 1\}$, where

$$R_g = \frac{p(\mathbf{K}'_l, G'_l | \text{rest})}{p(\mathbf{K}_l, G_l | \text{rest})} \frac{q(G_l | G'_l, z_l)}{q(G'_l | G_l, z_l)} \frac{J(\mathbf{K}'_l \rightarrow \phi'_l)}{J(\mathbf{K}_l \rightarrow \phi_l)} \frac{\pi_{m_l}(G'_l)}{\pi_{m_l}(G_l)} \frac{J(\phi_l \rightarrow \phi'_l)}{\frac{1}{\sigma_g \sqrt{2\pi}} \exp\left(-\frac{((\phi'_l)_{v_1, v_2} - (\phi_l)_{v_1, v_2})^2}{2\sigma_g^2}\right)}.$$

Since the free elements of ϕ'_l are the free elements of ϕ_l and $(\phi'_l)_{v_1, v_2}$, the Jacobian of the transformation from ϕ_l to ϕ'_l is equal to 1. Moreover $\det \mathbf{K}'_l = \det \mathbf{K}_l$ and $J(\mathbf{K}'_l \rightarrow \phi'_l) = (\phi_l)_{v_1, v_1} J(\mathbf{K}_l \rightarrow \phi_l)$. It follows that

$$R_g = \sigma_g \sqrt{2\pi} (\phi_l)_{v_1, v_1} z_l \frac{I_{G_l}(\delta_l, \mathbf{D}_l)}{I_{G'_l}(\delta_l, \mathbf{D}_l)} \frac{q(G_l | G'_l, z_l)}{q(G'_l | G_l, z_l)} \frac{\pi_{m_l}(G'_l)}{\pi_{m_l}(G_l)} \times \exp\left\{-\frac{1}{2} \langle \mathbf{K}'_l - \mathbf{K}_l, \mathbf{S}_l + z_l \mathbf{D}_l \rangle + \frac{((\phi'_l)_{v_1, v_2} - (\phi_l)_{v_1, v_2})^2}{2\sigma_g^2}\right\}.$$

Next we assume that G'_l is obtained by deleting the edge (v_1, v_2) from G_l . We have $\nu(G'_l) = \nu(G_l) \setminus \{(v_1, v_2)\}$, $\text{bd}_{G'_l}(v_1) = \text{bd}_{G_l}(v_1) \setminus \{v_2\}$ and $d_{v_1}^{G'_l} = d_{v_1}^{G_l} - 1$. We define an upper diagonal matrix ϕ'_l such that $(\phi'_l)_{v'_1, v'_2} = (\phi_l)_{v'_1, v'_2}$ for all $(v'_1, v'_2) \in \nu(G'_l)$. The bound elements of ϕ'_l are obtained by completion with respect to G'_l . The candidate precision matrix is $\mathbf{K}'_l = (\phi'_l)^T \phi'_l \in \mathbf{P}_{G'_l}$. Since the dimensionality of the parameter space decreases by 1, the acceptance probability of the update of (\mathbf{K}_l, G_l) to (\mathbf{K}'_l, G'_l) is $\min\{R'_g, 1\}$, where

$$R'_g = (\sigma_g \sqrt{2\pi} (\phi_l)_{v_1, v_1} z_l)^{-1} \frac{I_{G_l}(\delta_l, \mathbf{D}_l)}{I_{G'_l}(\delta_l, \mathbf{D}_l)} \frac{q(G_l | G'_l, z_l)}{q(G'_l | G_l, z_l)} \frac{\pi_{m_l}(G'_l)}{\pi_{m_l}(G_l)} \times \exp\left\{-\frac{1}{2} \langle \mathbf{K}'_l - \mathbf{K}_l, \mathbf{S}_l + z_l \mathbf{D}_l \rangle - \frac{((\phi'_l)_{v_1, v_2} - (\phi_l)_{v_1, v_2})^2}{2\sigma_g^2}\right\}.$$

5. MULTI-WAY GGMs WITH SEPARABLE MEAN PARAMETERS

So far we have discussed multi-way GGMs associated with array normal distributions with a $m_1 \times \dots \times m_L$ array mean parameter \mathbf{M} assumed to be zero. In some practical applications this assumption is too restrictive and \mathbf{M} needs to be explicitly accounted for. The observed samples $\mathcal{D} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\}$ grouped as an $m_1 \times \dots \times m_L \times n$ array are modeled as

$$(17) \quad \mathcal{D} = \mathbf{M} \circ \mathbf{1}_n + \mathbf{X}, \quad \mathbf{X} \sim \text{AN}_{L+1}(\mathbf{0}; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\}, \{n, \mathbf{K}_{L+1}\}), \quad \mathbf{K}_{L+1} = \mathbf{I}_n.$$

If the sample size n is small or if the observed samples are not independent and their dependence structure is represented by removing the constraint $\mathbf{K}_{L+1} = \mathbf{I}_n$, estimating $m = \prod_{l=1}^L m_l$ mean parameters is unrealistic. The matrix-variate normal models of [3] have separable mean parameters defined by row and column means, while [2] extends separable means for general array data. The L -dimensional mean array \mathbf{M} is written as a sum of distinct mean arrays associated with each dimension:

$$\mathbf{M} = \sum_{l=1}^L \mathbf{M}_l, \text{ where } \mathbf{M}_l = \mathbf{1}_{m_1} \circ \dots \circ \mathbf{1}_{m_{l-1}} \circ \boldsymbol{\mu}_l \circ \mathbf{1}_{m_{l+1}} \circ \dots \circ \mathbf{1}_{m_L}, 1 \leq l \leq L.$$

Here $\boldsymbol{\mu}_l \in \mathbb{R}^{m_l}$ represents the mean vector associated with dimension l of \mathbf{X} . This particular structure of the mean array \mathbf{M} implies the following marginal distribution for each element of the array of random effects \mathbf{X} :

$$X_{i_1 \dots i_L i_{L+1}} \sim \mathcal{N} \left(\sum_{l=1}^L (\boldsymbol{\mu}_l)_{i_l}, \prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i_l} \right).$$

Thus $\boldsymbol{\mu}_l$ can be interpreted as fixed effects associated with dimension l . The dependency between two different elements of \mathbf{X} is represented by their covariance

$$\text{Cov}(X_{i_1 \dots i_L i_{L+1}}, X_{i'_1 \dots i'_L i'_{L+1}}) = \prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i'_l}.$$

We remark that the individual mean arrays \mathbf{M}_l are not identified, but their sum \mathbf{M} is identified.

Bayesian estimation of \mathbf{M} proceeds by specifying independent priors $\boldsymbol{\mu}_l \sim \mathcal{N}_{m_l}(\boldsymbol{\mu}_l^0, \boldsymbol{\Omega}_l^{-1})$. To simplify the notations we take $n = 1$, hence the arrays in equation (17) have only L dimensions. We develop a Gibbs sampler in which each vector $\boldsymbol{\mu}_l$ is updated as follows. Denote $m_{-l} = \prod_{l' \neq l} m_{l'}$ and consider the l -mode matricizations $\mathcal{D}_{(l)}$, $\mathbf{M}_{(l)}$ and $\mathbf{X}_{(l)}$ of the arrays \mathcal{D} , \mathbf{M} and \mathbf{X} . From equation (17) it follows that the $m_l \times m_{-l}$ random matrix

$$\widetilde{\mathbf{X}}_{(l)} = \mathbf{X}_{(l)} - \sum_{l' \neq l} (\mathbf{M}_{l'})_{(l)}$$

follows a matrix-variate normal distribution with mean $(\mathbf{M}_l)_{(l)} = \boldsymbol{\mu}_l \mathbf{1}_{m_{-l}}^T$, row precision matrix \mathbf{K}_l and column precision matrix $\mathbf{K}_{-l} = \mathbf{K}_L \otimes \dots \otimes \mathbf{K}_{l+1} \otimes \mathbf{K}_{l-1} \otimes \dots \otimes \mathbf{K}_1$. It follows that $\boldsymbol{\mu}_l$ is updated by direct sampling from the multivariate normal $\mathcal{N}_{m_l}(\mathbf{m}_{\mu_l}, \mathbf{K}_{\mu_l}^{-1})$ where

$$\mathbf{K}_{\mu_l} = (\mathbf{1}_{m_{-l}}^T \mathbf{K}_{-l} \mathbf{1}_{m_{-l}}) \mathbf{K}_l + m_{-l} \boldsymbol{\Omega}_l, \quad \mathbf{m}_{\mu_l} = \mathbf{K}_{\mu_l}^{-1} [\mathbf{K}_l \widetilde{\mathbf{X}}_{(l)} \mathbf{K}_{-l} \mathbf{1}_{m_{-l}} + m_{-l} \boldsymbol{\Omega}_l \boldsymbol{\mu}_l^0].$$

6. AN APPLICATION TO SPATIOTEMPORAL CANCER MORTALITY SURVEILLANCE

In this section we construct a spatial hierarchical model for spatiotemporal cancer mortality surveillance based on the multi-way GGMs just developed. A relevant dataset could comprise counts $y_{i,j,t}$ for the number of deaths from cancer i in area j on year t , and can be seen as a three-dimensional array of size $m_C \times m_S \times m_T$. Our proposed model accounts for temporal and spatial dependence in mortality counts, as well as dependence across cancer types:

$$y_{i_C, i_S, i_T} \mid \theta_{i_C, i_S, i_T} \sim \text{Poi}(\exp\{\mu_{i_C} + \log(h_{i_S, i_T}) + \theta_{i_C, i_S, i_T}\}).$$

Here $i_C = 1, \dots, m_C$, $i_S = 1, \dots, m_S$, $i_T = 1, \dots, m_T$, h_{i_S, i_T} denotes the population in area i_S during year i_T , μ_{i_C} is the mean number of deaths due to cancer i_C over the whole period and all locations, and θ_{i_C, i_S, i_T} is a zero-mean random effect, which is assigned the prior:

$$\boldsymbol{\Theta} = (\theta_{i_C, i_S, i_T}) \sim \text{AN}_3(\mathbf{0}; \{m_C, \mathbf{K}_C\}, \{m_S, \mathbf{K}_S\}, \{m_T, \mathbf{K}_T\}).$$

The matrix \mathbf{K}_C models dependence across cancer types, \mathbf{K}_S accounts for spatial dependence across neighboring areas, and \mathbf{K}_T accounts for temporal dependence. Since we do not have prior information about dependence across cancer types, the prior for \mathbf{K}_C is specified hierarchically by setting

$$\mathbf{K}_C \mid G_C \sim \text{Wis}_{G_C}(\delta_C, \mathbf{I}_{m_C}), \quad p(G_C) \propto 1.$$

Thus $G_C \in \mathcal{G}_{m_C}$ defines the unknown graphical model of cancer types. For the spatial component, we follow the approach of [26] and use a GGM to specify a conditionally autoregressive prior

$$(z_S \mathbf{K}_S) \mid G_S \sim \text{Wis}_{G_S}(\delta_S, (\delta_S - 2)\mathbf{D}_S),$$

where $\mathbf{D}_S = (\mathbf{E}_W - \rho \mathbf{W})^{-1}$ and W is the adjacency matrix for the m_S areas, so that $W_{i_S^1, i_S^2} = 1$ if areas i_S^1 and i_S^2 share a common border, and $W_{i_S^1, i_S^2} = 0$ otherwise, and $\mathbf{E}_W = \text{diag}\{\mathbf{1}_{m_S}^T \mathbf{W}\}$. The graph G_S is fixed and given by the adjacency matrix \mathbf{W} . Furthermore, we assume that, *a priori*, there is a strong degree of positive spatial association, and choose a prior for spatial autocorrelation parameter ρ that gives higher probabilities to values close to 1 (see [33]):

$$\rho \sim \text{Uni}(\{0, 0.05, 0.1, \dots, 0.8, 0.82, \dots, 0.90, 0.91, \dots, 0.99\}).$$

For the temporal component, the prior for \mathbf{K}_T is set to

$$(z_T \mathbf{K}_T) \mid G_T \sim \text{Wis}_{G_T}(\delta_T, \mathbf{I}_{m_T}).$$

The graph G_T gives the temporal pattern of dependence and could be modeled in a manner similar to the graph G_C for cancer types. Instead of allowing G_T to be any graph with m_T vertices, we can constrain it to belong to a restricted set of graphs, for example, the graphs $G_T^{(1)}$, $G_T^{(2)}$, $G_T^{(3)}$ and $G_T^{(4)}$ with vertices $\{1, 2, \dots, m_T\}$ and edges $E_T^{(1)}$, $E_T^{(2)}$, $E_T^{(3)}$ and $E_T^{(4)}$, where

$$\begin{aligned} E_T^{(1)} &= \{(i_T - 1, i_T) : 2 \leq i_T \leq m_T\}, \\ E_T^{(2)} &= E_T^{(1)} \cup \{(i_T - 2, i_T) : 3 \leq i_T \leq m_T\}, \\ E_T^{(3)} &= E_T^{(2)} \cup \{(i_T - 3, i_T) : 4 \leq i_T \leq m_T\}, \\ E_T^{(4)} &= E_T^{(3)} \cup \{(i_T - 4, i_T) : 5 \leq i_T \leq m_T\}. \end{aligned}$$

These four graphs define AR(1), AR(2), AR(3) and AR(4) models. We set $\delta_C = \delta_S = \delta_T = 3$. We use a multivariate normal prior for the mean rates vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_{m_C})^T \sim \mathbf{N}_{m_C}(\boldsymbol{\mu}^0, \boldsymbol{\Omega}^{-1})$ where $\boldsymbol{\mu}^0 = \mu_0 \mathbf{1}_{m_C}$ and $\boldsymbol{\Omega} = \omega^{-2} \mathbf{I}_{m_C}$. We set μ_0 to be the median log incidence rate across all cancers, all areas and all time points, and ω to be twice the interquartile range of the raw log incidence rates.

The MCMC algorithm for this sparse multivariate spatiotemporal model involves iterative updates of the precision matrices \mathbf{K}_C , \mathbf{K}_S and \mathbf{K}_T as well as of the graph G_C as described in Section 4. The mean rates $\boldsymbol{\mu}$ are sampled as described in Section 5. Here the three-dimensional mean parameter array \mathbf{M} is equal with the mean array associated with the first dimension (cancers), while the mean arrays associated with the other two dimensions (space and time) are set to zero:

$$\mathbf{M} = \boldsymbol{\mu} \circ \mathbf{1}_{m_S} \circ \mathbf{1}_{m_T}.$$

We consider the centered random effects $\tilde{\boldsymbol{\Theta}} = \mathbf{M} + \boldsymbol{\Theta}$ which follows an array normal distribution with mean \mathbf{M} and precision matrices \mathbf{K}_C , \mathbf{K}_S , \mathbf{K}_T . We form $\tilde{\boldsymbol{\Theta}}_{(1)}$ — the 1-mode matricization of $\tilde{\boldsymbol{\Theta}}$. It follows that $\tilde{\boldsymbol{\Theta}} = \tilde{\boldsymbol{\Theta}}_{(1)}^T$ is a $(m_S m_T) \times m_C$ matrix that follows a matrix-variate normal distribution with mean $\mathbf{1}_{m_S m_T} \boldsymbol{\mu}^T$, row precision matrix $\tilde{\mathbf{K}}_R = \mathbf{K}_T \otimes \mathbf{K}_S$ and column precision matrix \mathbf{K}_C . We resample $\tilde{\boldsymbol{\Theta}}$ by sequentially updating each row vector $\tilde{\boldsymbol{\Theta}}_{i\Lambda}$, $i = 1, \dots, m_S m_T$. Conditional on the other rows of $\tilde{\boldsymbol{\Theta}}$, the distribution of $(\tilde{\boldsymbol{\Theta}}_{i\Lambda})^T$ with $i = i_S i_T$ ($1 \leq i_S \leq m_S$, $1 \leq i_T \leq m_T$) is multivariate normal with mean \mathbf{M}_i and precision matrix \mathbf{V}_i , where

$$\mathbf{M}_i = \boldsymbol{\mu} - \sum_{i'=1}^{m_S m_T} \frac{(\tilde{\mathbf{K}}_R)_{ii'}}{(\tilde{\mathbf{K}}_R)_{ii}} \left[(\tilde{\boldsymbol{\Theta}}_{i'\Lambda})^T - \boldsymbol{\mu} \right], \quad \mathbf{V}_i = (\tilde{\mathbf{K}}_R)_{ii} \mathbf{K}_C.$$

Thus the full conditional distribution of $\tilde{\boldsymbol{\Theta}}_{i\Lambda}$ is proportional with

$$\begin{aligned} & \prod_{i_C=1}^{m_C} \exp \{ y_{i_C, i_S, i_T} (\mu_{i_C} + \log(h_{i_S, i_T}) + \theta_{i_C, i_S, i_T}) - h_{i_S, i_T} \exp(\mu_{i_C} + \theta_{i_C, i_S, i_T}) \} \times \\ (18) \quad & \times \exp \left\{ -\frac{1}{2} \left[\tilde{\boldsymbol{\Theta}}_{i\Lambda} - (\mathbf{M}_i)^T \right] \mathbf{V}_i \left[(\tilde{\boldsymbol{\Theta}}_{i\Lambda})^T - \mathbf{M}_i \right] \right\}. \end{aligned}$$

We make use of a Metropolis-Hastings step to sample from to sample from (18). We consider a strictly positive precision parameter $\tilde{\sigma}$. For each $i_C = 1, \dots, m_C$, we update the i_C -th element of $\tilde{\Theta}_{i\Lambda}$ by sampling $\gamma \sim N(\tilde{\Theta}_{i,i_C}, \tilde{\sigma}^2)$. We define a candidate row vector $\tilde{\Theta}_{i\Lambda}^{new}$ by replacing $\tilde{\Theta}_{i,i_C}$ with γ in $\tilde{\Theta}_{i\Lambda}$. We update the current i -th row of $\tilde{\Theta}$ with $\tilde{\Theta}_{i\Lambda}^{new}$ with the Metropolis-Hastings acceptance probability corresponding with (18). Otherwise the i -th row of $\tilde{\Theta}$ remains unchanged.

7. DYNAMIC MULTI-WAY GGMs FOR ARRAY-VARIATE TIME SERIES

The cancer mortality surveillance application from Section 6 represented the time component as one of the dimensions of the three-dimensional array of observed counts. We give an extension of multi-way GGMs to array-variate time series $\mathbf{Y}_t, t = 1, 2, \dots, T$, where $\mathbf{Y}_t \in \mathbb{R}^{m_1 \times \dots \times m_L}$. Our framework generalizes the results from [14, 15] and [59] which assume vector ($L = 1$) or matrix-variate ($L = 2$) time series. We build on the standard specification of Bayesian dynamic linear models [61], and assume that \mathbf{Y}_t is modeled over time by

$$(19) \quad \mathbf{Y}_t = \mathbf{\Theta}_t \times_{L+1} \mathbf{F}_t^T + \mathbf{\Psi}_t, \quad \mathbf{\Psi}_t \sim \text{AN}_L(\mathbf{0}; \{m_1, v_t^{-1} \mathbf{K}_1\}, \{m_2, \mathbf{K}_2\}, \dots, \{m_L, \mathbf{K}_L\}),$$

$$(20) \quad \mathbf{\Theta}_t = \mathbf{\Theta}_{t-1} \times_{L+1} \mathbf{H}_t + \mathbf{\Gamma}_t, \quad \mathbf{\Gamma}_t \sim \text{AN}_{L+1}(\mathbf{0}; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\}, \{s, \mathbf{W}_t^{-1}\}),$$

where (a) $\mathbf{\Theta}_t \in \mathbb{R}^{m_1 \times \dots \times m_L \times s}$ is the state array at time t ; (b) $\mathbf{F}_t \in \mathbb{R}^s$ is a vector of known regressors at time t ; (c) $\mathbf{\Psi}_t \in \mathbb{R}^{m_1 \times \dots \times m_L}$ is the array of observational errors at time t ; (d) \mathbf{H}_t is a known $s \times s$ state evolution matrix at time t ; (e) $\mathbf{\Gamma}_t \in \mathbb{R}^{m_1 \times \dots \times m_L \times s}$ is the array of state evolution innovations at time t ; (f) \mathbf{W}_t is the $s \times s$ innovation covariance matrix at time t ; (g) $v_t > 0$ is a known scale factor at time t . Furthermore, the observational errors $\mathbf{\Psi}_t$ and the state evolution errors $\mathbf{\Gamma}_t$ follow zero-mean array normal distributions defined by $\mathbf{K}_1, \dots, \mathbf{K}_L$ and \mathbf{W}_t , and are assumed to be both independent over time element-wise and mutually independent as sequences of arrays.

The observation equation (19) and the evolution equation (20) translate into the following dynamic linear model for the univariate time series $(\mathbf{Y}_t)_{i_1 \dots i_L}, t = 1, 2, \dots, T$:

$$\begin{aligned} (\mathbf{Y}_t)_{i_1 \dots i_L} &= \mathbf{F}_t^T (\mathbf{\Theta}_t)_{i_1 \dots i_L \star} + (\mathbf{\Psi}_t)_{i_1 \dots i_L}, \quad (\mathbf{\Psi}_t)_{i_1 \dots i_L} \sim N\left(0, v_t \prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i_l}\right), \\ (\mathbf{\Theta}_t)_{i_1 \dots i_L \star} &= \mathbf{H}_t (\mathbf{\Theta}_{t-1})_{i_1 \dots i_L \star} + (\mathbf{\Gamma}_t)_{i_1 \dots i_L \star}, \quad (\mathbf{\Gamma}_t)_{i_1 \dots i_L \star} \sim N_s\left(\mathbf{0}, \prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i_l} \mathbf{W}_t\right), \end{aligned}$$

where $(\mathbf{\Theta}_t)_{i_1 \dots i_L \star} = ((\mathbf{\Theta}_t)_{i_1 \dots i_L 1}, \dots, (\mathbf{\Theta}_t)_{i_1 \dots i_L s})^T$, while $(\mathbf{\Theta}_{t-1})_{i_1 \dots i_L \star}$ and $(\mathbf{\Gamma}_t)_{i_1 \dots i_L \star}$ are defined in a similar manner. The components \mathbf{F}_t , \mathbf{H}_t and \mathbf{W}_t are the same for all univariate time series, but the state parameters $(\mathbf{\Theta}_t)_{i_1 \dots i_L \star}$ as well as their scales of measurement defined by $\prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i_l}$ could be different across series. The cross-sectional dependence structure across individual time series at time t is induced by $\mathbf{K}_1, \dots, \mathbf{K}_L$ and \mathbf{W}_t :

$$\text{Cov}\left((v_t)_{i_1 \dots i_L}, (v_t)_{i'_1 \dots i'_L}\right) = v_t \prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i'_l}, \quad \text{Cov}\left((\mathbf{\Gamma}_t)_{i_1 \dots i_L \star}, (\mathbf{\Gamma}_t)_{i'_1 \dots i'_L \star}\right) = \prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i'_l} \mathbf{W}_t.$$

For example, if $\prod_{l=1}^L (\mathbf{K}_l^{-1})_{i_l i'_l}$ is large in absolute value, the univariate time series $(\mathbf{Y}_t)_{i_1 \dots i_L}$ and $(\mathbf{Y}_t)_{i'_1 \dots i'_L}$ exhibit significant dependence in the variation of their observational errors and state vectors. Appropriate choices for the matrix sequence $\mathbf{W}_t, t = 1, 2, \dots, T$, arise from the discount factors discussed in [61] as exemplified, among others, in [59]. The scale factors v_t can be set to 1, but other suitable values can be employed as needed.

The following result extends Theorem 1 of [59] to array-variate time series.

Theorem 1. *Let \mathcal{D}_0 be the prior information and denote by $\mathcal{D}_t = \{\mathbf{Y}_t, \mathcal{D}_{t-1}\}$ the information available at time $t = 1, 2, \dots, T$. We assume to have specified precision matrices $\mathbf{K}_1, \dots, \mathbf{K}_L$, the matrix sequence $\mathbf{W}_t, t = 1, 2, \dots, T$, as well as an initial prior for the state array at time 0,*

$$(\mathbf{\Theta}_0 | \mathcal{D}_0) \sim \text{AN}_{L+1}(\mathbf{M}_0; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\}, \{s, \mathbf{C}_0^{-1}\}),$$

where $\mathbf{M}_0 \in \mathbb{R}^{m_1 \times \dots \times m_L \times s}$ and \mathbf{C}_0 is an $s \times s$ covariance matrix. For every $t = 1, 2, \dots, T$, the following distributional results hold:

(i) posterior at $t - 1$:

$$(\boldsymbol{\Theta}_{t-1} \mid \mathcal{D}_{t-1}) \sim \text{AN}_{L+1}(\mathbf{M}_{t-1}; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\}, \{s, \mathbf{C}_{t-1}^{-1}\}),$$

(ii) prior at t :

$$(\boldsymbol{\Theta}_t \mid \mathcal{D}_{t-1}) \sim \text{AN}_{L+1}(\mathbf{a}_t; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\}, \{s, \mathbf{R}_t^{-1}\}),$$

where $\mathbf{a}_t = \mathbf{M}_{t-1} \times_{L+1} \mathbf{H}_t$ and $\mathbf{R}_t = \mathbf{H}_t \mathbf{C}_{t-1} \mathbf{H}_t^T + \mathbf{W}_t$.

(iii) one-step forecast at $t - 1$:

$$(\mathbf{Y}_t \mid \mathcal{D}_{t-1}) \sim \text{AN}_L(\mathbf{f}_t; \{m_1, q_t^{-1} \mathbf{K}_1\}, \{m_2, \mathbf{K}_2\}, \dots, \{m_L, \mathbf{K}_L\}),$$

where $\mathbf{f}_t = \mathbf{M}_{t-1} \times_{L+1} (\mathbf{F}_t^T \mathbf{H}_t) = \mathbf{a}_t \times_{L+1} \mathbf{F}_t^T$ and $q_t = \mathbf{F}_t^T \mathbf{R}_t \mathbf{F}_t + v_t$.

(iv) posterior at t :

$$(\boldsymbol{\Theta}_t \mid \mathcal{D}_t) \sim \text{AN}_{L+1}(\mathbf{M}_t; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\}, \{s, \mathbf{C}_t^{-1}\}),$$

where $\mathbf{M}_t = \mathbf{a}_t + \mathbf{e}_t \times_{L+1} \mathbf{A}_t$, $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t^T q_t$. Here $\mathbf{A}_t = q_t^{-1} \mathbf{R}_t \mathbf{F}_t$ and $\mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t$.

The proof of Theorem 1 is straightforward. We write equations (19) and (20) in matrix form:

$$(21) \quad (\mathbf{Y}_t)_{(L+1)} = \mathbf{F}_t^T (\boldsymbol{\Theta}_t)_{(L+1)} + (\boldsymbol{\Psi}_t)_{(L+1)}, \quad (\boldsymbol{\Psi}_t)_{(L+1)} \sim \mathbf{N}_m(\mathbf{0}, v_t \mathbf{K}^{-1}),$$

$$(22) \quad (\boldsymbol{\Theta}_t)_{(L+1)} = \mathbf{H}_t (\boldsymbol{\Theta}_{t-1})_{(L+1)} + (\boldsymbol{\Psi}_t)_{(L+1)}, \quad (\boldsymbol{\Psi}_t)_{(L+1)} \sim \text{AN}_2(\mathbf{0}; \{s, \mathbf{W}_t^{-1}\}, \{m, \mathbf{K}\}),$$

where $m = \prod_{l=1}^L m_l$ and \mathbf{K} is given in equation (6). The normal theory results laid out in [61] apply directly to the dynamic linear model specified by equations (21) and (22). The predictive distributions relevant for forecasting and retrospective sampling for array-variate time series can be derived from the corresponding predictive distributions for vector data.

We complete the definition and prior specification for the dynamic multi-way GGMs with independent G-Wishart priors from equation (10) for the precision matrices $\mathbf{K}_1, \dots, \mathbf{K}_L$ and their corresponding auxiliary variables z_2, \dots, z_L . The graphs G_1, \dots, G_L associated with the G-Wishart priors receive independent priors $\pi_{m_l}(G_l)$ on \mathcal{G}_{m_l} , $l = 1, \dots, L$. Posterior inference in this framework can be achieved with the following MCMC algorithm that sequentially performs the following steps:

(A) *Resampling the precision matrices, graphs and auxiliary variables.* By marginalizing over the state arrays $\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_L$, we obtain the marginal likelihood [14, 15]:

$$\mathbf{p}(\mathbf{Y}_1, \dots, \mathbf{Y}_T \mid \mathbf{K}_1, G_1, (\mathbf{K}_l, G_l, z_l)_{l=2}^L) = \prod_{t=1}^T \mathbf{p}(\mathbf{Y}_t \mid \mathcal{D}_{t-1}, \mathbf{K}_1, G_1, (\mathbf{K}_l, G_l, z_l)_{l=2}^L).$$

The one-step forecast distribution (iii) from Theorem 1 implies that

$$(q_t^{-1/2}(\mathbf{Y}_t - \mathbf{f}_t) \mid \mathcal{D}_{t-1}) \sim \text{AN}_L(\mathbf{0}; \{m_1, \mathbf{K}_1\}, \{m_2, \mathbf{K}_2\}, \dots, \{m_L, \mathbf{K}_L\}).$$

We use the filtering equations from Theorem 1 to produce the centered and scaled array data $\bar{\mathcal{D}} = \{q_t^{-1/2}(\mathbf{Y}_t - \mathbf{f}_t) : t = 1, \dots, T\}$. Since the elements of $\bar{\mathcal{D}}$ are independent and identically distributed, we update each precision matrix \mathbf{K}_l , graph G_l and auxiliary variable z_l as described in Section 4 based on $\bar{\mathcal{D}}$.

(B) *Resampling the state arrays.* We employ the forward filtering backward algorithm (FFBS) proposed by [12, 30]. Given the current sampled precision matrices, we start by sampling $\boldsymbol{\Theta}_T$ given \mathcal{D}_T from the posterior distribution given in (iv) of Theorem 1. Then, for $t = T - 1, T - 2, \dots, 0$, we sample $\boldsymbol{\Theta}_t$ given \mathcal{D}_T and $\boldsymbol{\Theta}_{t+1}$ from the array normal distribution

$$\text{AN}_{L+1}(\mathbf{M}_t^*; \{m_1, \mathbf{K}_1\}, \dots, \{m_L, \mathbf{K}_L\}, \{s, (\mathbf{C}_t^*)^{-1}\}),$$

where

$$\mathbf{M}_t^* = \mathbf{M}_t + (\boldsymbol{\Theta}_{t+1} - \mathbf{a}_{t+1}) \times_{L+1} (\mathbf{C}_t \mathbf{G}_{t+1}^T \mathbf{R}_{t+1}^{-1}).$$

8. DISCUSSION

Recent advances in data collection techniques have allowed the creation of high-dimensional public health datasets that monitor the incidence of many diseases across several areas, time points and additional ecological sociodemographic groupings [28]. Jointly modeling the disease risk associated with each resulting cell count (i.e., a particular disease at a particular time point in a particular region given a particular combination of risk factors) is desirable since it takes into consideration interaction patterns that arise within each dimension or across dimensions. By aggregating data across time, key epidemiological issues related to the evolution of the risk patterns across time might not be given an appropriate answer [1]. The spatial structure of geographical regions must also be properly accounted for [6, 9]. Furthermore, since diseases are potentially related and share risk factors, it is critical that individual models should not be developed for each disease [33, 56]. Rich, flexible classes of models that capture the joint variation of disease risk in the actual observed data without requiring the aggregation across one or more dimensions will be the fundamental aim of our proposed work related to disease mapping. Multi-way GGMs can be used in Bayesian hierarchical models that produce estimates of disease risk by borrowing strength across time, areas and the other dimensions. Due to the likely presence of small counts in many cells, the degree of smoothing will be controlled through a wide range of parameters that could be constrained to zero according to pre-defined interaction structures (e.g., the neighborhood structure of the areas) or by graphs that received the most support given the data.

We generalize the models from Section 6, and let Y be the L -dimensional array of observed disease counts indexed by cells $\{(i_1, \dots, i_L) : 1 \leq i_l \leq m_l\}$. We assume that the count random variable $Y_{i_1 \dots i_L}$ associated with cell (i_1, \dots, i_L) follows a distribution from an exponential family (e.g., Poisson or binomial) with mean parameter $\theta_{i_1 \dots i_L}$, i.e.

$$(23) \quad Y_{i_1 \dots i_L} \mid \theta_{i_1 \dots i_L} \stackrel{iid}{\sim} H(\theta_{i_1 \dots i_L}), \text{ for } 1 \leq i_l \leq m_l, 1 \leq l \leq L.$$

We assume that the cell counts Y are conditionally independent given the L -dimensional array of parameters $\theta = \{\theta_{i_1 \dots i_L} : 1 \leq i_l \leq m_l\}$. Furthermore, given a certain link function $g(\cdot)$ (e.g., $\log(\cdot)$), the parameters θ follow a joint model

$$(24) \quad g(\theta_{i_1 \dots i_L}) = \nu_{i_1 \dots i_L} + X_{i_1 \dots i_L},$$

where $\nu_{i_1 \dots i_L}$ is a known offset, while $X = \{X_{i_1 \dots i_L} : 1 \leq i_l \leq m_l\}$ is an array of zero-centered random effects. Equation (24) can subsequently include explanatory ecological covariates as needed. The multi-way GGMs are employed in the context of non-Gaussian data as joint distribution for the array of random effects X . Thus X is assumed to follow the flexible joint distributions, and Each dimension of the data is represented as a GGM in a particular dimension of the random effects X .

This framework accommodates many types of interactions by restricting the set of graphs that are allowed to represent the dependency patterns of the corresponding dimensions. For example, if dimension l' of Z represents time, then the graphs associated with this dimension could be constrained to represent an autoregressive model $AR(q)$, where $q = 1, 2, 3, \dots$ – see Section 6. Temporal dependence can also be modeled with the dynamic multi-way GGMs from Section 7. If dimension l'' represents spatial dependence, one could constrain the space of graphs for dimension l'' to consist of only one graph with edges defined by areas that are neighbors of each other in the spirit of [6, 16]. As opposed to a modeling framework based on GMRFs, we can allow uncertainty around this neighborhood graph in which case we let the space of graphs for dimension l'' to include graphs that are obtained by adding or deleting one, two or more edges from the neighborhood graph. This expansion of the set of spatial graphs is consistent with the hypothesis that interaction occurs not only between areas that are close to each other or share a border, but also between more distant areas. We can also allow all possible graphs to be associated with dimension l'' and examine the graphs that receive the highest posterior probabilities. Such graphs can be further compared with the neighborhood graph to see whether the spatial dependency patterns in observed data are actually consistent with the geographical neighborhoods.

To gain further insight on the flexibility of our modeling approach, we examine the case in which a two-dimensional array $Y = \{Y_{i_1 i_2}\}$ is observed with the first dimension associated with m_1 diseases and the second dimension associated with m_2 areas. Under the framework of [33, 11] the matrix of counts

Y is modeled with a hierarchical Poisson model with random effects distributed as a multivariate CAR (MCAR) model [46]:

$$(25) \quad X' \sim N_{m_1 m_2}(0, [K_1 \otimes (E_W - \rho W)]^{-1}).$$

This structure of the random effects assumes separability of the association structure among diseases from the spatial structure [55]. The spatial autocorrelation parameter ρ is the only parameter that controls the strength of spatial dependencies, while the precision matrix K_1 is not subject to any additional constraints on its elements. In our framework, the random effects X' follow a matrix-variate GGM prior obtained by taking $L = 2$ in equation (6). The same separability of the association structure is assumed, but the precision matrices K_1 and K_2 follow G-Wishart hyper-priors as in equation (10). The GGMs associated with the diseases are allowed to vary across all possible graphs with m_1 vertices, while the GGMs for the spatial structure can be modeled as we described earlier in this section.

Until recently, the application of GGMs with a G-Wishart prior for the precision matrix in large scale Bayesian hierarchical models has been hindered by computational difficulties. For decomposable graphs, the normalizing constant of the G-Wishart distribution is calculated with formulas [50, 4], and a direct sampler from this distribution existed for several years [13]. But similar results did not exist for non-decomposable graphs. Fortunately, new methodological developments give formulas for the calculation of the G-Wishart distribution for arbitrary graphs [54], and also a direct sampler for arbitrary graphs [41]. With these key results, the MCMC sampler developed in Section 4 can be significantly improved in its efficiency. The reversible jump algorithm that allows updates in the structure of the graphs associated the dimensions of a multi-way GGM can be subsequently refined to another transdimensional graph updating algorithm which bypasses the calculation of any normalizing constants of the G-Wishart distribution based on the double reversible jump algorithm of [58, 41]. Moreover, the G-Wishart distribution can be replaced altogether in the specification of priors for Bayesian hierarchical spatial models with the graphical lasso prior of [57]. The application of these new theoretical results to spatial health data is a very intense area of research.

REFERENCES

- [1] J. J. Abellan, S. Richardson, and N. Best. Use of space-time models to investigate the stability of patterns of disease. *Environmental Health Perspectives*, 116:1111–1119, 2008.
- [2] G. I. Allen. Comment on article by Hoff. *Bayesian Analysis*, 6:197–202, 2011.
- [3] G. I. Allen and R. Tibshirani. Inference with transposable data: Modeling the effects of row and column correlations. arXiv:1004.0209v1 [stat.ME], 2010.
- [4] A. Atay-Kayis and H. Massam. A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, 92:317–35, 2005.
- [5] J. Barnard, R. McCulloch, and X. Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statist. Sinica*, 10:1281–311, 2000.
- [6] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.
- [7] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.
- [8] J. Besag and C. Kooperberg. On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746, 1995.
- [9] J. Besag, J. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59, 1991.
- [10] P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36:199–227, 2008.
- [11] B. P. Carlin and S. Banerjee. Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors, *Bayesian Statistics 7*, pages 45–63. Oxford: Oxford University Press, 2003.
- [12] C. K. Carter and R. Kohn. Gibbs sampling for state space models. *Biometrika*, 81:541–553, 1994.
- [13] C. M. Carvalho, H. Massam, and M. West. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94:647–659, 2007.
- [14] C. M. Carvalho and M. West. Dynamic matrix-variate graphical models. *Bayesian Analysis*, 2:69–98, 2007.
- [15] C. M. Carvalho and M. West. Dynamic matrix-variate graphical models—a synopsis. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, editors, *Bayesian Statistics VIII*, pages 585–590. Oxford: Oxford University Press, 2007.
- [16] D. G. Clayton and J. M. Kaldor. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681, 1987.
- [17] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1973.

- [18] M. Daniels and R. Kass. Nonconjugate Bayesian estimation of covariance matrices. *J. Am. Statist. Assoc.*, 94:1254–63, 1999.
- [19] A. P. Dawid. Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika*, 68:265–274, 1981.
- [20] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21:1253–1278, 2000.
- [21] P. Dellaportas, P. Giudici, and G. Roberts. Bayesian inference for nondecomposable graphical Gaussian models. *Sankhyā*, 65:43–55, 2003.
- [22] A. P. Dempster. Covariance selection. *Biometrics*, 28:157–75, 1972.
- [23] P. Diaconnis and D. Ylvisaker. Conjugate priors for exponential families. *Ann. Statist.*, 7:269–81, 1979.
- [24] A. Dobra, C. Hans, B. Jones, J. R. Nevins, G. Yao, and M. West. Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, 90:196–212, 2004.
- [25] A. Dobra and A. Lenkoski. Copula Gaussian graphical models and their application to modeling functional disability data. *Annals of Applied Statistics*, 5:969–993, 2011.
- [26] A. Dobra, A. Lenkoski, and A. Rodriguez. Bayesian inference for non-decomposable general gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association*, 106:1418–1433, 2011.
- [27] M. Drton and M. D. Perlman. Model selection for Gaussian concentration graphs. *Biometrika*, 91:591–602, 2004.
- [28] P. Elliott, J. Wakefield, N. Best, and D. Briggs. *Spatial Epidemiology: Methods and Applications*. Oxford University Press, 2001.
- [29] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 3:432–441, 2008.
- [30] S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.
- [31] A. T. Galecki. General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics - Theory and Methods*, 23:3105–3119, 1994.
- [32] A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp. *Handbook of Spatial Statistics*. CRC Press: Taylor & Francis Group, 2010.
- [33] A. E. Gelfand and P. Vounatsou. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4:11–25, 2003.
- [34] P. Giudici and P. J. Green. Decomposable graphical Gaussian model determination. *Biometrika*, 86:785–801, 1999.
- [35] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(711–732), 1995.
- [36] P. D. Hoff. Separable covariance arrays via the Tucker product, with applications to multivariate relational data. Technical report, Department of Statistics, University of Washington, 2010.
- [37] B. Jones, C. Carvalho, A. Dobra, C. Hans, C. Carter, and M. West. Experiments in stochastic computation for high-dimensional graphical models. *Statist. Sci.*, 20:388–400, 2005.
- [38] R. Kass and A. E. Raftery. Bayes factors. *J. Am. Statist. Assoc.*, 90:773–95, 1995.
- [39] T. G. Kolda. Multilinear operators for higher-order decompositions. Technical report, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, 2006.
- [40] S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.
- [41] A. Lenkoski. A direct sampler for G-Wishart variates. *Stat.*, 2:119–128, 2013.
- [42] A. Lenkoski and A. Dobra. Computational aspects related to inference in Gaussian graphical models with the G-wishart prior. *Journal of Computational and Graphical Statistics*, 20:140–157, 2011.
- [43] T. Leonard and J. S. J. Hsu. Bayesian inference for a covariance matrix. *Ann. Statist.*, 20:1669–96, 1992.
- [44] G. Letac and H. Massam. Wishart distributions for decomposable graphs. *Ann. Statist.*, 35:1278–323, 2007.
- [45] J. C. Liechty, M. W. Liechty, and P. Müller. Bayesian correlation estimation. *Biometrika*, 91:1–14, 2004.
- [46] K. V. Mardia. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, 24:265–284, 1988.
- [47] N. Meinshausen and P. Bühlmann. High-dimensional graphs with the Lasso. *Ann. Statist.*, 34:1436–62, 2006.
- [48] R. A. Olshen and B. Rajaratnam. Successive normalization of rectangular arrays. *Annals of Statistics*, 38:1638–1664, 2010.
- [49] B. Rajaratnam, H. Massam, and C. M. Carvalho. Flexible covariance estimation in graphical Gaussian models. *Ann. Statist.*, 36:2818–2849, 2008.
- [50] A. Roverato. Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Statist.*, 29:391–411, 2002.
- [51] H. Rue and L. Knorr-Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall / CRC, 2005.
- [52] J. G. Scott and C. M. Carvalho. Feature-inclusion stochastic search for Gaussian graphical models. *J. Comput. Graph. Statist.*, 17:790–808, 2008.
- [53] M. Smith and R. Kohn. Bayesian parsimonious covariance matrix estimation for longitudinal data. *J. Am. Statist. Assoc.*, 87:1141–53, 2002.
- [54] C. Uhler, A. Lenkoski, and D. Richards. Exact formulas for the normalizing constants of Wishart distributions for graphical models. <http://arxiv.org/abs/1406.4901>, 2014.

- [55] L. Waller and B. P. Carlin. Disease mapping. In A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, editors, *Handbook of Spatial Statistics*, pages 217–243. CRC Press: Taylor & Francis Group, 2010.
- [56] F. Wang and M. M. Wall. Generalized common factor spatial model. *Biostatistics*, 4:569–582, 2003.
- [57] H. Wang. The Bayesian graphical lasso and efficient posterior computation. *Bayesian Analysis*, 7:771–790, 2012.
- [58] H. Wang and S. Z. Li. Efficient gaussian graphical model determination under G-wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198, 2012.
- [59] H. Wang and M. West. Bayesian analysis of matrix normal graphical models. *Biometrika*, 96:821–834, 2009.
- [60] N. Wermuth. Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics*, 32:95–108, 1976.
- [61] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer - Verlag, New York, second edition, 1997.
- [62] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, 1990.
- [63] F. Wong, C. K. Carter, and R. Kohn. Efficient estimation of covariance selection models. *Biometrika*, 90:809–830, 2003.
- [64] R. Yang and J. O. Berger. Estimation of a covariance matrix using the reference prior. *Ann. Statist.*, 22:1195–211, 1994.
- [65] M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94:19–35, 2007.

DEPARTMENT OF STATISTICS, DEPARTMENT OF BIOBEHAVIORAL NURSING AND HEALTH SYSTEMS, CENTER FOR STATISTICS AND THE SOCIAL SCIENCES AND CENTER FOR STUDIES IN DEMOGRAPHY AND ECOLOGY, UNIVERSITY OF WASHINGTON, BOX 354322, SEATTLE, WA 98195

E-mail address: adobra@uw.edu